



UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
09/930,014	08/15/2001	Steve Goddard	UNL3058.1	1924

7590 12/02/2004

MICHAEL J. THOMAS
HARNES, DICKEY & PIERCE
7700 BONHOMME AVENUE
SUITE 400
ST. LOUIS, MI 63105

EXAMINER

PWU, JEFFREY C

ART UNIT	PAPER NUMBER
----------	--------------

2143

DATE MAILED: 12/02/2004

Please find below and/or attached an Office communication concerning this application or proceeding.

Office Action Summary

Application No.

09/930,014

Applicant(s)

GODDARD, STEVE

Examiner

Jeffrey Pwu

Art Unit

2143

-- The MAILING DATE of this communication appears on the cover sheet with the correspondence address --

Period for Reply

A SHORTENED STATUTORY PERIOD FOR REPLY IS SET TO EXPIRE 3 MONTH(S) FROM THE MAILING DATE OF THIS COMMUNICATION.

- Extensions of time may be available under the provisions of 37 CFR 1.136(a). In no event, however, may a reply be timely filed after SIX (6) MONTHS from the mailing date of this communication.
- If the period for reply specified above is less than thirty (30) days, a reply within the statutory minimum of thirty (30) days will be considered timely.
- If NO period for reply is specified above, the maximum statutory period will apply and will expire SIX (6) MONTHS from the mailing date of this communication.
- Failure to reply within the set or extended period for reply will, by statute, cause the application to become ABANDONED (35 U.S.C. § 133). Any reply received by the Office later than three months after the mailing date of this communication, even if timely filed, may reduce any earned patent term adjustment. See 37 CFR 1.704(b).

Status

- 1) ☐ Responsive to communication(s) filed on ____.
- 2a) ☐ This action is **FINAL**. 2b) ☒ This action is non-final.
- 3) ☐ Since this application is in condition for allowance except for formal matters, prosecution as to the merits is closed in accordance with the practice under *Ex parte Quayle*, 1935 C.D. 11, 453 O.G. 213.

Disposition of Claims

- 4) ☒ Claim(s) 1-38 is/are pending in the application.
- 4a) Of the above claim(s) ____ is/are withdrawn from consideration.
- 5) ☐ Claim(s) ____ is/are allowed.
- 6) ☒ Claim(s) 1-38 is/are rejected.
- 7) ☐ Claim(s) ____ is/are objected to.
- 8) ☐ Claim(s) ____ are subject to restriction and/or election requirement.

Application Papers

- 9) ☐ The specification is objected to by the Examiner.
- 10) ☐ The drawing(s) filed on ____ is/are: a) ☐ accepted or b) ☐ objected to by the Examiner.
Applicant may not request that any objection to the drawing(s) be held in abeyance. See 37 CFR 1.85(a).
Replacement drawing sheet(s) including the correction is required if the drawing(s) is objected to. See 37 CFR 1.121(d).
- 11) ☐ The oath or declaration is objected to by the Examiner. Note the attached Office Action or form PTO-152.

Priority under 35 U.S.C. § 119

- 12) ☐ Acknowledgment is made of a claim for foreign priority under 35 U.S.C. § 119(a)-(d) or (f).
- a) ☐ All b) ☐ Some * c) ☐ None of:
- ☐ Certified copies of the priority documents have been received.
 - ☐ Certified copies of the priority documents have been received in Application No. ____.
 - ☐ Copies of the certified copies of the priority documents have been received in this National Stage application from the International Bureau (PCT Rule 17.2(a)).

* See the attached detailed Office action for a list of the certified copies not received.

Attachment(s)

- 1) ☒ Notice of References Cited (PTO-892)
- 2) ☐ Notice of Draftsperson's Patent Drawing Review (PTO-948)
- 3) ☒ Information Disclosure Statement(s) (PTO-1449 or PTO/SB/08)
Paper No(s)/Mail Date 3/31/04
- 4) ☐ Interview Summary (PTO-413)
Paper No(s)/Mail Date. ____.
- 5) ☐ Notice of Informal Patent Application (PTO-152)
- 6) ☐ Other: ____.

DETAILED ACTION

Title

1. The title of the invention is not descriptive. A new title is required that is clearly indicative of the invention to which the claims are directed.

Claim Objections

2. Claim 2 is objected to because of the following informalities:

“2. The server of claim 2...” should be changed to ‘2. The server of claim 1...’.

Appropriate correction is required.

Claim Rejections - 35 USC § 102

3. The following is a quotation of the appropriate paragraphs of 35 U.S.C. 102 that form the basis for the rejections under this section made in this Office action:

A person shall be entitled to a patent unless –

(b) the invention was patented or described in a printed publication in this or a foreign country or in public use or on sale in this country, more than one year prior to the date of application for patent in the United States.

4. Claims 1-38 are rejected under 35 U.S.C. 102(b) as being anticipated by Dias et al. (U.S. 6,490,615).

Dias et al. disclose claims:

1. A server for providing data to clients, the server comprising:
a dispatcher (160) having a queue for storing requests received from clients; and at least one back-end server (180);
wherein the dispatcher stores in the queue one or more of the requests received from clients when the back-end server is unavailable to process said one or more requests (col.5, lines 10-25);

Art Unit: 2143

wherein the dispatcher retrieves said one or more requests from the queue for forwarding to the back-end server when the back-end server becomes available to process said one or more requests (col.5, lines 10-25); and wherein the dispatcher determines whether the back-end server is available to process said one or more requests by comparing a number of connections concurrently supported by the back-end server to a maximum number of concurrent connections that the back-end server is permitted to support, the maximum number being less than a maximum number of connections which the back-end server is capable of supporting concurrently (col.5, lines 10-25; "When client 110, 120, or 130 issues a request to retrieve an object (i.e. a target object), the request is routed and transmitted by network dispatcher 160 to a first cache of cache array 170. A first-cache selected by network dispatcher 160 may or may not correspond to the requested object. In other words, the requested object may or may not be assigned to the first cache. If the first cache selected by network dispatcher 160 corresponds to the requested object (i.e. it is the primary owner of the requested object) and the requested object is stored in the first cache, then the first cache may service the request. If the first cache selected by network dispatcher 160 corresponds to the requested object (i.e. it is the primary owner of the requested object) and the requested object is not stored on the first cache, then the first cache may retrieve the requested object from an appropriate server 182, 184, 186, or 188 of server cluster 180.")

2. The server of claim 1 wherein the dispatcher is configured to monitor a performance of the back-end server, to define the maximum number of concurrent connections that the back-end server is permitted to support, and to dynamically adjust the maximum number in response to the monitored performance (abstract).

3. The server of claim 1 wherein the server is a cluster-based server comprising a plurality of back-end servers, the dispatcher is configured to store in the queue said one or more requests when none of the back-end servers are available to process said one or more requests, and the dispatcher is further configured to retrieve said one or more requests from the queue for forwarding to one of the back-end servers when said one of the back-end servers becomes available to process said one or more requests (180, fig.1; col.3, line 55-col.5, line 25).

4. The server of claim 1 wherein the server is a Web server (abstract)

Art Unit: 2143

5. The server of claim 1 wherein the dispatcher and the back-end server are implementing using COTS hardware (it is inherent to use any available commercially off-the shelf hardware).

6. The server of claim 1 wherein the dispatcher comprises a first computer device, the back-end server comprises a second computer device, and the first and second computer devices are configured to communicate with one another over a computer network (see figs. 1a, 2, 3, 4, 5, 6, 7, and 8).

7. The server of claim 1 wherein the dispatcher is an OSI layer 7 dispatcher and said requests are data requests (col.2, line65-col.3, line 8).

8. The server of claim 7 wherein the dispatcher implements a simplified TCP/IP protocol in user-space (col.2, line65-col.3, line 8).

9. The server of claim 1 wherein the dispatcher is an OSI layer 4 dispatcher and said requests are connection requests (col.3, lines 20-37).

10. A computer-readable medium having computer executable instructions for performing the method of claim 1 (abstract).

11. A method for controlled server loading, the method comprising the steps of:
defining a maximum number of concurrent connections that a server is permitted to support (abstract, col.9, lines 1-37);
limiting a number of concurrent connections supported by the server to the maximum number;
monitoring the server's performance while it supports the concurrent connections; and
dynamically adjusting the maximum number as a function of the server's performance to thereby control a performance factor for the server (figs. 1a, 2, 3, 4, 5, 6, 7, and 8).

12. The method of claim 11 wherein the defining step includes defining the maximum number to be less than a maximum number of connections which the server is capable of supporting concurrently (col.5, lines 10-25; "When client 110, 120, or 130 issues a request to retrieve an object (i.e. a target object), the request is routed and transmitted by network dispatcher 160 to a first cache of cache array 170. A first-cache selected by network dispatcher 160 may or may not correspond to the requested object. In other words, the requested object may or may not be assigned to the first' cache. If the first cache selected by network dispatcher 160 corresponds to the requested object (i.e. it is the primary owner of the requested object) and the requested object is stored in the first cache, then the first cache may service the request. If the first cache selected

Art Unit: 2143

by network dispatcher 160 corresponds to the requested object (i.e. it is the primary owner of the requested object) and the requested object is not stored on the first cache, then the first cache may retrieve the requested object from an appropriate server 182, 184, 186, or 188 of server cluster 180.”)

13. The method of claim 11 wherein the concurrent connections are connections between the server and clients (figs. 1a, 2, 3, 4, 5, 6, 7, and 8).

14. The method of claim 11 wherein the concurrent connections are connections between the server and a dispatcher (figs. 1a, 2, 3, 4, 5, 6, 7, and 8).

15. The method of claim 11 wherein the server is a backend server in a cluster-based server having a dispatcher, and the dynamically adjusting step includes dynamically adjusting the maximum number of concurrent connections that can be established between the back-end server and the dispatcher (col.5, lines 10-25).

16. The method of claim 15 wherein each concurrent connection is a persistent connection over which data requests from multiple clients can be sent by the dispatcher to the back-end server (figs. 1a, 2, 3, 4, 5, 6, 7, and 8).

17. The method of claim 11 wherein the dynamically adjusting step includes dynamically adjusting the maximum number in response to the monitoring step such that the server operates at or above a minimum performance level (col.10, lines 55-68).

18. The method of claim 17 wherein the monitoring step includes monitoring the server's performance level in terms of a performance metric selected from the group consisting of request rate, average response time, maximum response time and server throughput (col.10, line 38-68)

Claims 19-38 are similarly rejected as in claims 1-18:

19. A method for controlled server loading, the method comprising the steps of:
receiving a plurality of data requests from clients;
forwarding a number of the data requests to a server for processing; and
storing at least one of the data requests until the server completes processing at least one of the forwarded data requests. (col.5, lines 10-25)

Art Unit: 2143

20. The method of claim 19 further comprising the steps of retrieving the stored data request after the server completes processing at least one of the forwarded data requests, and forwarding the retrieved data request to the server for processing.
21. The method of claim 19 wherein the storing step includes storing a plurality of the data requests, the method further comprising the step of retrieving one of the stored data requests and forwarding the retrieved one of the data requests to the server for processing each time the server completes processing one of the forwarded data requests.
22. The method of claim 21 wherein the retrieving step includes retrieving the stored data requests on a FIFO basis.
23. The method of claim 19 wherein the data requests are HTTP requests.
24. The method of claim 19 wherein the receiving, forwarding and storing steps are performed by a single computer device having at least one processor.
25. The method of claim 24 wherein the single computer device comprises the server.
26. The method of claim 19 wherein the storing step is performed by a dispatcher and includes storing at least one of the data requests until the dispatcher receives a response from the server to at least one of the forwarded data requests.
27. A method for controlled server loading, the method comprising the steps of:
defining a maximum number of data requests that a server is permitted to process concurrently;
monitoring the server's performance; and
dynamically adjusting the maximum number in response to the monitoring step to thereby adjust the server's performance.
28. The method of claim 27 wherein the monitoring step includes monitoring the server's performance in terms of a performance metric selected from the group consisting of request rate, average response time, maximum response time, and server throughput.
29. The method of claim 27 further comprising the steps of receiving a plurality of data requests from clients, forwarding some of the data requests to the server for processing, and storing at least one of the data requests until the server completes processing one of the forwarded data requests.

Art Unit: 2143

30. The method of claim 27 wherein the defining step includes defining a maximum number of connections that can be supported concurrently by the server and limiting the number of data requests that can be pending on each connection.

31. The method of claim 30 wherein the defining step includes limiting the number of data requests that can be pending on each connection to one.

32. A method for controlled loading of a cluster-based server, the cluster-based server including a dispatcher and a plurality of back-end servers, the method comprising the steps of receiving at the dispatcher a plurality of data requests from clients; forwarding a plurality of the data requests to each of the back-end servers for processing; and storing at the dispatcher at least one of the data requests until one of the back-end servers completes processing one of the forwarded data requests.

33. The method of claim 32 wherein the storing step includes storing a plurality of the data requests and the forwarding step includes forwarding one of the stored data requests to one of the back-end servers each time one of the back-end servers completes processing one of the forwarded data requests.

34. The method of claim 32 wherein the cluster-based server is an L7/3 server.

35. A method for controlled loading of a cluster-based server, the cluster-based server including a dispatcher and a plurality of back-end servers, the method comprising the steps of: defining, for each back-end server, a maximum number of data requests that can be processed concurrently; monitoring the performance of each back-end server; and dynamically adjusting the maximum number for at least one of the back-end servers in response to the monitoring step to thereby adjust the performance of the cluster-based server.

36. The method of claim 35 wherein the dynamically adjusting step includes dynamically adjusting the maximum number for each back-end server.

37. The method of claim 35 wherein the dynamically adjusting step includes dynamically adjusting the maximum number for said one of the back-end servers as a function of that back-end server's performance.

38. The method of claim 35 further comprising the steps of receiving a plurality of data requests from clients,

Art Unit: 2143

forwarding some of the data requests to the back-end servers for processing, and storing at least one of the data requests until one of the back-end servers completes processing one of the forwarded data requests.

5. Any inquiry concerning this communication or earlier communications from the examiner should be directed to Jeffrey Pwu whose telephone number is 571 272-6798. If attempts to reach the examiner by telephone are unsuccessful, the examiner's supervisor, David Wiley can be reached on 571 272-3923. The fax phone number for the organization where this application or proceeding is assigned is 703-872-9306.

Information regarding the status of an application may be obtained from the Patent Application Information Retrieval (PAIR) system. Status information for published applications may be obtained from either Private PAIR or Public PAIR. Status information for unpublished applications is available through Private PAIR only. For more information about the PAIR system, see <http://pair-direct.uspto.gov>. Should you have questions on access to the Private PAIR system, contact the Electronic Business Center (EBC) at 866-217-9197 (toll-free).



Sunday, November 21, 2004

JEFFREY PWU
PRIMARY EXAMINER